On the Relative Value of Averages derived from different numbers of Observations. By WILLIAM A. GUY, M.B., Cantab, *Fellow of the Royal College of Physicians; Professor of Forensic Medicine, King's College; Physician to King's College Hospital; Honorary Secretary of the Statistical Society.*

[Read before the Statistical Society of London, 16th April, 1849.]

THE subject of this communication has been usually treated as a branch of the mathematics; and few, if any, attempts have been made to illustrate it by means of observation. The labour which this mode of illustration must necessarily entail has probably disinclined men from its adoption. On the other hand, abstract reasoning possesses the twofold advantage of brevity and certainty. The mathematician sees at once that any attempt to establish broad principles, or to construct formulæ, by the aid of observation alone, must necessarily fail, from the vast number of facts which would be required before even a starting-point for calculation could be reached. To take a familiar example:—the chances of throwing doublets with the dice may be very easily calculated; but days, weeks, or even months, might be spent in a vain attempt to deduce these chances from actual observation. So, also, with Vital Statistics. It would not be possible to accumulate a sufficient number of facts in illustration of the age at death of any single class of persons to enable us to determine how many of such facts would suffice to establish a true average. There is only one obvious method by which such a result could be accomplished; and that is, by placing, side by side, two averages based upon the same number of facts, and adding an equal number of facts to either total, till the difference between the averages fell to zero, and continued to stand at that point through several successive additions. Even if we were to content ourselves with a rougher approximation, it could only be reached by a very laborious series of calculations, which might not, after all, repay the pains bestowed upon them. It is not, therefore, with any hope of solving the difficult question—How many observations are necessary to obtain a true average? that the following facts are adduced; but simply to furnish an illustration, imperfect though it be, of the variable results obtained by actual observation on a limited scale. The facts themselves, on which the averages are founded, happen to have been collected in such a manner as to fit them for the use to which they are now put; and having spent some time in the calculations, I am unwilling that the results should be lost. They may perhaps possess some immediate interest, and be put hereafter to some useful purpose.

The facts to which I refer consist of the ages, at death, of the members of the several ranks and professions, and have already supplied the materials of a series of communications to the Statistical Society. Three collections of such facts are available for my present purpose: the first, relating to the duration of life of the English aristocracy; the second, to that of kings; the third, to that of the large mixed class, whose deaths are recorded in the obituaries of the Annual

Register, consisting of almost every man of note in the higher and middle ranks of society during about three quarters of a century.

It may be well to premise that, in throwing these several classes of facts into groups, care has been taken to avoid everything bordering on selection. Those facts which came first to hand are placed first, and those which followed in the order of collection take the second place in the tables. The same caution has been observed in adding group to group. The first use which I propose to make of the facts is to arrange them in averages of 25, 50, 75, 100, and so on up to 1000, in two parallel columns, with a third column of differences, with a view of ascertaining the rate and degree of approximation of the two series.

TABLE I.

Average Age at Death.

No. of Facts.	1st Series.	2nd Series.	Difference.
25	60·96	61·24	0·28
50	59·46	60·66	1·20
75	61·30	61·20	0·10
100	61·78	61·98	0·20
125	61·10	61·42	0·32
150	61·90	60·61	1·29
175	62·44	60·67	1·77
200	61·34	60·85	0·49
225	61·07	60·44	0·63
250	60·56	59·93	0·63
275	59·95	59·63	0·32
300	59·77	60·10	0·33
325	59·80	59·88	0·88
350	59·84	60·14	0·30
375	59·93	59·85	0·08
400	59·35	59·98	0·63
425	59·42	60·00	0·58
450	59·38	60·23	0·85
475	59·37	60·43	1·06
500	59·42	60·66	1·24
525	59·26	60·57	1·31
550	59·14	60·74	1·60
575	59·34	60·62	1·28
600	59·48	60·75	1·27
625	59·53	60·56	1·03
650	59·53	60·26	0·73
675	59·58	60·28	0·70
700	59·58	60·31	0·73
725	59·80	60·30	0·50
750	59·79	60·49	0·70
775	59·72	60·38	0·66
800	59·84	60·67	0·83
825	59·96	60·52	0·56
850	59·77	60·62	0·85
875	59·74	60·61	0·87
900	59·78	60·47	0·69
925	59·87	60·56	0·69
950	59·87	60·46	0·59
975	59·90	60·38	0·48
1,000	59·75	60·38	0·63

Table I. consists of two such columns, each headed by an average of 25 facts, deduced from the first 50 ages at death, extracted from the peerage. Each group of 25 facts becomes 50 in the second line, 75 in the third line, 100 in the fourth line, each group increasing by 25 at a time until the last average in each column is based on 1000 facts. The facts successively added were taken, like the first hundred facts, from the peerage, and then from the baronetage, in the exact order in which they occur.

This table will at least serve to show the hopelessness of the attempt to discover by the mere accumulation of observations, the number of facts which may be necessary to furnish a true average. It will be seen that the two columns of figures continue to exhibit as wide divergences and as marked fluctuations between the average values derived from a large as from a small number of facts; that the first difference is below the average of all the differences: and that, with a single exception, the averages derived from 75 facts show a closer approximation than any other figures in the two columns. The superiority of a large to a small number of facts shows itself, however, in the steadiness of the results (if we strike off the decimals) in all the averages obtained from more than 250 facts in the first column, and from upwards of 400 in the second column. It will also be observed that the difference between the highest and lowest average in each column of figures is very small. In the first column the greatest average is 62·44, the least 59·14, and the difference 3·30; in the second column, the highest average is 61·98, the lowest 59·63, and the difference 2·35. It would appear, then, if this table were taken as a guide, that averages drawn from even a small number of facts do not lead to those extreme inaccuracies to which they are generally supposed to be liable. But, as conclusions based upon a single fact, or a single collection of facts, must always be viewed with suspicion, it is very desirable to extend our investigation so as to embrace at least a second collection of figures of the same order. My inquiries into the duration of life among sovereigns have supplied me with the means of effecting this on a small scale, in Table II.

In this table it will be observed that the first average of the first column falls short of the first average of the second column by nearly five years, and that all the subsequent averages of the two columns exhibit a much closer approximation. In like manner, the difference between the highest and lowest averages of either column is more considerable, being $54·84 - 48·88 = 5·96$, in the one case, and $56·91 - 52·92 = 3·99$ in the other. Both columns also exhibit a greater amount of fluctuation than those of Table I. The natural inference to be drawn from this circumstance is, that the slight difference between the first averages of the two columns in the first table was a mere coincidence, and that the more uniform and steady character of that table is due to such coincidence.

It is obvious, therefore, that tables constructed upon this principle would be liable to lead careless reasoners into error, and that a different arrangement of the elementary facts is necessary to the right understanding of the relative value of different numbers of facts as data for reasoning.

TABLE II.

Average Age at Death.			
No.of Fac s.	1st Series.	2nd Series.	Difference.
25	48·88	53·72	4·84
50	51·24	52·92	1·68
75	53·79	54·23	0·44
100	54·84	55·56	0·72
125	54·61	56·08	1·47
150	54·38	56·91	2·53
175	54·17	56·21	2·04
200	54·34	55·99	1·65
225	54·50	56·19	1·69
250	54·36	56·14	1·78
275	54·09	55·93	1·84
300	53·82	55·80	1·98
325	54·01	55·91	1·90
350	54·40	55·85	1·45
375	54·28	56·08	1·80
400	54·14	56·20	2·06
425	54·35	55·79	1·44
450	54·37	55·57	1·20
475	53·90	55·26	1·36
500	53·82	55·28	1·36
525	53·86	55·05	1·19
550	53·92	54·77	0·85

The following tables, of which the first is founded on the average ages, at death, of the peerage and baronetage, and the second, on the average ages, at death, of the members of the several ranks and professions, as deduced from the "Annual Register," will be found to comprise most of the elements of an instructive comparison. The facts in both instances were taken without selection, in the order in which they stood in my original papers. The several facts were first arranged in a line in groups of 25 each, two successive groups of 25 were then formed into groups of 50, the groups of 50 into groups of 100, and so on, till the last totals in the tables were obtained.

TABLE III.

No. of Facts.	No. of Groups.	Average Age at Death.		
		Maximum.	Minimum.	Range.
25	64	69·40	50·64	18·76
50	32	66·44	55·20	11·24
100	16	63·70	56·85	6·85
200	8	62·38	57·61	4·77
400	4	61·10	58·24	2·86
800	2	60·84	59·67	1·17
1,600	1	60·25		....

TABLE IV.

No. of Facts.	No. of Groups.	Average Age at Death.		
		Maximum.	Minimum.	Range.
50	128	84·44	56·78	27·66
100	64	76·24	58·25	17·99
200	32	73·54	61·50	12·04
400	16	69·78	63·51	6·27
800	8	68·67	65·07	3·60
1,600	4	67·93	64·84	3·09
3,200	2	66·38	65·82	0·56
6,400	1	66·10		....

Each of these tables exhibits, in a very striking manner, the wide difference which may exist between averages deduced from small numbers of facts. The first table, founded on facts of a very uniform character, namely, the ages, at death, of what may be termed the one class of the English aristocracy, exhibits, for averages based on the small number of 25 observations, a possible difference between the highest and lowest average of nearly 19 years, being considerably above one-fourth of the highest average. In the second table, which is formed of units differing more widely from each other, inasmuch as they represent the ages at death of several classes of the community, the difference between the highest and lowest average of 50 observations is, in round numbers, 28 years, or nearly one-third of the highest average. If we take the true average in the first table to be 60 years (the mean of 1600 observations), then the greatest average exhibits an error of more than 9 years, and the least average an error of about the same amount. If, again, in the second table, we assume the true average to be 66 (the average of 6,400 observations), we may have an error in excess to the extent of 18 years, and an error in defect to the extent of 10 years. But when these averages are used for purposes of comparison, it is obvious that the possible error may amount to the sum of the two errors involved in the two averages respectively. Suppose, for example, that we wish to ascertain the relative longevity of the members of the English aristocracy and of the entire upper and middle class, and we proceed to determine the question by averages founded on 50 observations, it might happen that the average for the first-named class was the minimum 55·20, and for the last-named class the maximum 84·44. The difference between these two numbers is 29·24. But on the assumption that the true averages are 60·25 and 66·38, the true difference will be only 6·13. So that the possible error is no less than 29·24 — 6·13, or 23·11.

It will be seen that the limits of possible error diminish rapidly with an increase of observations. Thus, if we take the first table, and express the range, or difference, in round numbers, the limit of error in excess or defect diminishes as half the respective numbers 19, 11, 7, 5, 3, 1. In the case of the second table, the limits of error will be as the half of the several round numbers 28, 18, 12, 6, 4, 3, 1. We look in vain for a numerical law of approximation, unless, indeed, the round

numbers 1, 3, 5, 7, in the last four lines of the first table, exhibiting, as they do, an arithmetical progression with the number 2 as the common difference, may be regarded as an indication of such a law. We could only adopt this supposition by the somewhat arbitrary assumption that the law does not begin to display itself till a rude approximation to a true average has been obtained by summing up 100 observations. The great irregularity of the second table, moreover, would seem to forbid this view of the case.

There is one defect in these tables which may serve to explain the absence of any decided approach to regularity in the figures which represent the limits of variation; namely, the circumstance that the number of the groups diminishes as the number of facts increases. For instance, the wide range of 19 years, which appears in the first line of the first table, is the difference between the highest and lowest averages of no less than 64 groups of 25 facts, while the limited range of one year is obtained from only two groups of 800 facts each. If any approximation to a numerical law of increase or decrease is to be looked for as the result of observation, it is clearly not reasonable to expect it except from a comparison of the same number of groups. And although it is probable that the chance of the discovery of such a law in such a manner would be small, unless the groups were not merely equal in number, but also numerous, I have thought that some traces of such a law might possibly be discovered in a collection of facts in which the groups should be equal, though limited in number. The eighth Annual Report of the Registrar-General furnishes the materials for a comparison of this kind, by presenting us with the number of male and female births for the several counties and registration districts of England, during each of the six years 1839-44. From this return I have selected, without previous calculation of the results, and with an eye solely to the number of the facts, certain districts and counties; and, having calculated for each of the six years, the number of male births which would have happened in one million of births, have given in separate columns the average number of births recorded in the year, the greatest and least number of male births in a million, and the range or difference between the two extremes. The results are embodied in Table V.

This table, in common with Tables III. and IV., serves to enforce the little dependence to be placed on small numbers of facts, but it is still less favourable to the discovery, by observation, of any numerical law of approximation. The figures which express the range or difference between the highest and the lowest number of male births, present no approach to regularity; and there is every reason to believe that the fluctuations are due, not to the variable proportion of male and female births, but merely, or chiefly, to the larger or smaller number of facts. If we endeavour to arrive at some numerical law by throwing the several returns into larger groups, we are equally unfortunate. If, for instance, we throw the 28 separate returns into four groups of seven each, strike an average of each group, and place the last return for all England by itself, we obtain, as the differences between the highest and lowest number of male births in one million, the numbers 3,028, 6,584, 9,812, 24,648, and 59,655, in which it is quite hopeless to attempt to trace any law of increase or decrease.

We are equally baffled if, allowing the first and last return to stand by themselves, we distribute the remaining 27 returns into three groups of 9 each. The five differences in this case are as follows: 3,028, 6,663, 12,605, 47,163, 107,016.

TABLE V.

Name of District.	Number of Births.	Maximum.	Minimum.	Range.
Salisbury	271	543,798	436,782	107,016
Canterbury	368	556,180	482,500	73,680
Penkridge	436	554,113	493,333	60,780
Winchester	615	545,031	492,228	52,803
Rutlandshire	722	527,246	466,942	60,304
Staines and Uxbridge	858	532,184	498,195	33,998
Northampton	1,047	524,281	495,274	29,007
St. George in the East	1,515	518,619	480,771	37,848
Alresford and Petersfield	1,612	541,256	500,000	41,256
Huntingdonshire	2,032	519,824	485,030	34,794
Middlesex (part of)	3,854	517,260	496,915	20,345
Cambridgeshire	5,940	522,690	507,188	15,502
Derbyshire	7,760	517,746	508,166	9,582
Suffolk	10,011	519,205	505,996	13,209
Gloucestershire	11,894	522,582	497,945	24,637
Kent	14,256	516,997	506,651	10,346
Staffordshire	15,771	516,439	507,767	8,672
South Wales	16,727	516,753	512,006	4,747
Northern Counties	27,916	517,735	511,330	6,405
Eastern Counties	32,212	514,620	509,765	4,855
North Midland Counties	36,451	519,801	510,778	9,023
South Midland Counties	37,764	515,455	507,205	8,250
South-Eastern Counties	44,290	519,268	508,978	10,290
South-Western Counties	52,656	514,210	509,368	4,842
York	53,825	515,244	510,453	4,791
Metropolis	59,422	513,812	509,491	4,321
Western Counties	61,903	515,808	507,642	8,166
North-Western Counties	76,721	515,985	510,557	5,428
England	515,478	514,809	511,781	3,028

Being thus completely baffled in my attempt to discover, by means of such observations as were most readily available, a numerical law of approximation, expressive of the relative value of averages founded upon different numbers of facts, I proceeded to compare the results of observation, in the matter of male and female births, with the liability to error, as derived from well known and generally received mathematical formulæ, of the several averages founded upon few or many facts.

But before I proceed to institute this comparison between the results of observation and the figures derived from the formulæ of the mathematician, I would revert, for a moment, to the tables already brought forward, with a view of throwing light upon a question of the greatest practical importance, namely, to what extent are we justified in employing averages founded upon small numbers of observations?

One of the leading results just established by the tables, is the wideness of the limits which separate the highest and lowest averages

derived from equal small groups of facts. For instance, the averages drawn from 64 groups of 25 ages at death of peers and baronets, ranged between a minimum of 51 and a maximum of 69 years, leaving a difference of no less than 18 years, the average of 1,600 observations being about 60 years. In like manner 128 groups, containing each 50 ages at death, of the mixed upper and middle classes, gave a maximum of 84, a minimum of 57, and a range of no less than 27 years, the average of 6,400 facts being 66 years. The extremes approached, and the range contracted rapidly as the numbers of observations in each group increased.

Now this wide divergence of the extreme values derived from equal small groups of facts would seem at first sight to be absolutely fatal to the use of such small collections of facts for statistical purposes. But before we adopt this conclusion, it would be well to examine the several averages which lie between the two extremes, with a view to determine whether they are distributed equally or not; what are the chances that any average taken at hazard will approximate to the true mean, or to the extremes; and, generally, whether the chance of encountering, in the majority of instances, a near approach to the true value, may not be such as to warrant the employment of even small collections of facts, if not as demonstrative evidence, at least as valuable probabilities?

With a view of throwing some light upon this very important question, I have prepared two tables, constructed in the same manner, and showing, for the several groups of facts, the precise number of averages corresponding to each age lying between the two extremes. Each average is referred to the round numbers to which it is nearest. The average of all the facts is distinguished by a larger type.

TABLE VI.

Facts taken from the Peerage and Baronetage.

Average Age.	25 Facts, 72 Groups.	50 Facts, 36 Groups.	100 Facts, 18 Groups.	200 Facts, 9 Groups.	400 Facts, 4 Groups.	800 Facts, 2 Groups.
69	1	....	....	....	....	....
68	0	....	....	....	....	....
67	0	....	....	....	....	....
66	3	1	....	....	....	....
65	4	0	....	....	....	....
64	9	2	1	....	....	....
63	3	3	1	....	....	....
62	3	5	3	3	....	....
61	9	6	3	0	3	1
60	9	4	3	2	0	1
59	8	5	3	3	0	....
58	6	5	1	1	1	....
57	5	1	3	....	....	....
56	6	3	....	....	....	....
55	1	1	....	....	....	....
54	3	....	....	....	....	....
53	1	....	....	....	....	....
52	0	....	....	....	....	....
51	1	....	....	....	....	....

TABLE VII.

Facts taken from the Annual Register.

Average Age.	50 Facts, 128 Groups.	100 Facts, 64 Groups.	200 Facts, 32 Groups.	400 Facts, 16 Groups.	800 Facts, 8 Groups.	1,600 Facts, 4 Groups.	3,200 Facts, 2 Groups.
84	1	....	....	....	....	....	....
83	0	....	....	....	....	....	....
82	0	....	....	....	....	....	....
81	0	....	....	....	....	....	....
80	0	....	....	....	....	....	....
79	0	....	....	....	....	....	....
78	0	....	....	....	....	....	....
77	0	....	....	....	....	....	....
76	1	1	....	....	....	....	....
75	1	0	....	....	....	....	....
74	4	1	1	....	....	....	....
73	2	1	0	....	....	....	....
72	2	1	0	....	....	....	....
71	5	0	0	....	....	....	....
70	4	4	0	1	....	....	....
69	7	3	2	0	1	....	....
68	13	5	3	1	0	1	....
67	17	9	4	4	1	0	....
66	20	13	14	4	2	1	2
65	5	13	2	4	3	2	....
64	16	6	2	2	1	....	....
63	11	1	2	....	....	....	....
62	10	4	1	....	....	....	....
61	3	1	1	....	...	....	....
60	3	0	....	....	....	....	....
59	0	0	....	....	....	....	....
58	0	1	....	....	....	....	....
57	3	....	....	....	....	....	....

These tables supply a ready answer to the questions just propounded: they show at a glance, what *à priori* reasoning would suggest, that the individual averages are few in number in the direction of the extremes, and numerous as we approach the true average. In the first table, for instance, in no less than 9 cases out of 72, or one-eighth of the whole number, the averages of 25 facts are the same as the true average, namely 60; while, in the second table, 20 out of 128 groups of 50 facts, or little less than one-sixth of the whole number, yield the true average, namely, 66. The remaining columns of the two tables present similar results; but those of the second table are the most striking. Thus, of the 64 groups of 100 facts, 13, or about one-fifth, of the 32 groups of 200 facts, 14, or nearly one-half, of the 16 groups of 400 facts, 4, or one-fourth, and of the groups of 800 and 1600 facts, the like proportion are found to coincide with the true average derived from all the facts.

If, again, we take the instances in which the averages exceed, and fall short of the true average by only one year (which may be regarded as a very near approach to accuracy), we find that, in the first table, 26 out of 72 groups of 25 facts, 15 out of 36 groups of 50 facts, 9 out of 18 groups of 100 facts, 5 out of 9 groups of 200 facts, and 3 out of 4 groups of 400 facts, or upwards of one-third, little less than one-

half, exactly one-half, more than one-half, and three-fourths respectively, answer to this description. So also in the second table: 42 in 128 groups of 50, 35 in 64 groups of 100, 20 in 32 groups of 200, 12 in 16 groups of 400, 6 in 8 groups of 800, and 3 out of 4 groups of 1600, correspond with the average of all the facts. There is, therefore, no room for doubt that the chances in favour of an average, approaching very closely to the true average, greatly overbalance those in favour of an average approaching either extreme, even when the number of facts from which such average is calculated is inconsiderable.

The progressive and steady increase or decrease in series of averages of 25 or 50 facts (as in my observations on the frequency of the pulse in the two sexes at different ages), may be adduced in confirmation of the result of the foregoing tables, and as a sufficient reason for not rejecting conclusions based upon a small number of observations*.

The materials of the foregoing tables are the ages at death of members of the aristocracy, and of the combined upper and middle classes. I now propose to extend my inquiry, so as to embrace a different order of facts belonging to that class of investigations in which two alternative events, such as death or recovery from disease, the birth of a male or female child, &c., are in question. In order that the element of more or fewer facts might have full play, and be subject to the least possible disturbance, I have chosen a very simple alternative event, namely, the attendance of a male or female patient among the out-patients of an hospital. The facts have been carefully extracted from the physicians' out-patient book of the King's College hospital, in which are entered the names of all patients of both sexes above ten years, with exceptions which it is not necessary to specify. The male and female attendances were extracted from the books by fifties, in the exact order in which they were entered. The number of males in each fifty patients was then written down in a vertical line, two adjoining fifties were bracketed together to form a group of 100, two adjoining hundreds to form a group of 200, and so on till the grand total of 6,400 was obtained.

I shall assume that the average of these 6,400 facts form the true average, and as such, shall use it for a standard of comparison.

The most superficial examination of my tables convinced me that the fluctuations in this new order of facts were not less than those which I had encountered in the averages based upon the ages at death of different classes of the community.

Table VIII., which is a counterpart of Tables III. and IV., exhibits the per centage attendances of males, in maxima and minima, as obtained from groups of 50, 100, 200, 400, 800, 1,600, and 3,200 facts.

This table exhibits the same rapid but irregular approximations of the extreme values which characterize Tables III. and IV., and the same absence of any obvious numerical law of approximation. The per centage attendance of males which for 50 facts and 128 averages has a range equal to the average of all the groups, namely, a range of 40, exhibits for the two averages of 3,200 facts a range of less

* In the case of the female pulse, averages of 25 facts showed a regular and progressive decrease for every period of 7 years from birth to the 56th year, and a progressive increase from that age to the end of life. Averages of the same number of facts exhibited for the male pulse a similar but less steady decrease and increase.

than one-fifth of an unit. Even the four groups of 1,600 facts display a range of less than one and a-half: the rate of approximation to a true average is, therefore, very rapid.

TABLE VIII.

Number of Facts.	Number of Groups.	Attendance of Males in 100 Attendances.		
		Maximum.	Minimum.	Range.
50	128	64	24	40
100	64	53	29	24
200	32	46	$34\frac{1}{2}$	$11\frac{1}{2}$
400	16	$43\frac{3}{4}$	36	$7\frac{3}{4}$
800	8	$41\frac{1}{4}$	38	$3\frac{1}{4}$
1,600	4	$40\frac{7}{16}$	$39\frac{1}{8}$	$1\frac{5}{16}$
3,200	2	$39\frac{25}{32}$	$39\frac{31}{32}$	$\frac{3}{16}$
6,400	1	$39\frac{7}{8}$		

The probability of encountering a true average, or one equal to the average of all the facts, will also appear to be similar to that displayed in Tables VI. and VII. There is the same tendency in the averages to arrange themselves about the true mean, rather than in the neighbourhood of the extremes. This will fully appear in Table IX., in which, in consequence of the necessity of reducing all the averages to per centage proportions, the numbers in the column of 50 facts, being the figures extracted from the hospital books multiplied by 2, will be found opposite to even numbers. The figures in the other columns are the round numbers nearest to the quotients of the totals divided by 2, 4, 8, 16, and 32, respectively.

Having thus reproduced for a new order of facts the tables founded upon the ages at death of the aristocracy, and of the combined upper and middle classes, with a view of confirming the results arrived at by the use of these data, I proceed to institute the promised comparison between the results of actual observation and the calculations of the mathematician.

The materials which I have selected for this purpose are the numbers of male and female births which took place in several counties and registration districts of England and Wales during each of the six years from 1839 to 1844 inclusive. The comparison in question is embodied in Table X., of which the first column consists of numerals referring to the counties and registration districts specified in the foot-note; the second column, of the number of births registered in the several counties or districts on an average of the six years; the third column, of the mean number of male births in one million occurring in the six years; the fourth column, of the greatest and least number of male births in the six years, with the range, or difference between them; the fifth column, of the greatest and least number of male births which would have taken place in the same six years, on the supposition that the male and female births are really equal in number, and that the observed difference between the male and female births is an error of observation; the sixth column, of the error in excess or defect attaching to the number of facts upon that supposition; the seventh

column, of the greatest and least number of male births in one million, which would result from the addition and substraction of the error due to the number of facts to and from the average male births in one million in the same six years; and the eighth and last column, of the error in excess or defect to which the respective numbers of facts are liable. The formulæ by which the figures in the last four columns were calculated are given in the foot-notes.

TABLE IX.

Average No. of attendances of Males.	50 Facts.	100 Facts.	200 Facts.	400 Facts.	800 Facts.	1,600 Facts.	3,200 Facts.
64	1	....	....	....	....	....	....
63	....	....	....	....	....	....	....
62	....	....	....	....	....	....	....
61	....	....	....	....	....	....	....
60	1	....	....	....	....	....	....
59	....	....	....	....	....	....	....
58	....	....	....	....	....	....	....
57	....	....	....	....	....	....	....
56	2	....	....	....	....	....	....
55	....	....	....	....	....	....	....
54	1	....	....	....	....	....	....
53	....	1	....	....	....	....	....
52	3	....	....	....	....	....	....
51	....	....	....	....	....	....	....
50	1	....	....	....	....	....	....
49	....	1	....	....	....	....	....
48	8	2	....	....	....	....	....
47	....	....	....	....	....	....	....
46	11	1	1	....	....	....	....
45	....	2	2	....	....	....	....
44	13	7	1	....	....	....	....
43	....	5	2	1	....	....	....
42	15	6	1	1	....	....	....
41	....	5	5	4	2	....	....
40	15	5	3	5	3	3	2
39	....	4	3	1	2	1	....
38	14	5	7	2	1	....	....
37	....	5	3	1	....	....	....
36	8	6	2	1	....	....	....
35	....	1	1	....	....	....	....
34	14	3	1	....	....	....	....
33	....	2	....	....	....	....	....
32	9	2	....	....	....	....	....
31	....	....	....	....	....	....	....
30	5	....	....	....	....	....	....
29	....	1	....	....	....	....	....
28	5	....	....	....	....	....	....
27	....	....	....	....	....	....	....
26	1	....	....	....	....	....	....
25	....	....	....	....	....	....	....
24	2	....	....	....	....	....	....

The results of the comparison instituted in this table are interesting and instructive. On the supposition that the actual equality of male and female births in one district (No. 21,) in one of the six years, and the near approach to equality in other instances, had left us in doubt

whether the inequality in other years and districts might not be explained by the errors attaching to a comparatively small number of observations, the mathematician would direct us to compare columns 4 and 5, with a view to a solution of the difficulty. The figures in the fifth column are calculated upon the assumption, that the male and female births are naturally equal; but that in consequence of the limited number of facts embraced by the several returns, that equality is destroyed, and the proportion of male to female births is subject to great apparent fluctuation. If, then, this assumption of equality be false, the maxima of male births in the several districts ought uniformly to exceed the maxima obtained by calculation. In other words, in order to disprove the theory of equality, the maximum established by observation ought, in each return, to surpass the limits of possible error due to the numerical insufficiency of the facts.

But a reference to the table will show that it is only in nineteen out of twenty-nine returns that the highest number of male births obtained in each of the six years, 1839-44, does exceed the maximum obtained by calculation. In the rest of the returns, being ten in number, the use of the calculations in column 5 would at least leave us in doubt, as the observed maxima are less than the maxima derived from calculation. In these ten cases, therefore, though the returns display an apparent excess of male births, that excess falling short of the maxima derived from calculation, we should remain in doubt whether the observed excess of male births was in the nature of things, or merely an error of observation. To make this subject more easy of comprehension, I will take two cases in point—the returns for all England, marked 1, and the returns for Huntingdonshire, marked 20. On referring to column 5, it will be seen that, on the supposition of a perfect equality between the male and female births, 515,478 recorded births might yield any number between 500,624 and 499,376 male births in a million. Now, every one of the returns for the six years, 1839-44, greatly exceeds the maximum 500,624; it may, therefore, be fairly assumed that the male and female births are not equal in number, but that, in the nature of things, there is an excess of male births. On the other hand, it appears by column 5, that, on the same supposition of a perfect equality of male and female births, we might encounter in 2,032 births recorded in the county of Huntingdon any number of male births in a million between 531,369 and 468,631. But every one of the six returns from Huntingdonshire falls short of the maximum 531,369. Hence we may infer that at least there is the greatest possible doubt whether the male and female births in that county may not be equal. Now, there is good reason to believe that in England, at least, the rule of an excess of male births is subject to no exception. It obtains uniformly in the first fourteen districts of the table, from which the returns are most numerous; it fails only where the facts are comparatively few in number; and it is obviously in the highest degree improbable that out of twenty-nine returns taken without selection from several hundreds, the first fourteen should be governed by one rule, and the last fifteen by another. From this table, then, it may be fairly inferred that the formulæ of the mathematician are not applicable as tests to the results of observations founded on comparatively small numbers of facts.

A comparison of the actual maxima and minima established by observation in the six years, 1839-44, with the maxima and minima calculated by the formulæ $\frac{m}{\mu} \pm 2\sqrt{\frac{2mn}{\mu^3}}$ shows, as might be anticipated, a greater divergence between the extremes derived from calculation than between those established by observation within so limited a period. It is only in the single case of all England, where the number of facts exceeds half a million, that the extremes obtained by calculation fall within those established by observation.

On reviewing the tables contained in this communication to the Society, the following propositions may be put forward as fully warranted by them:

1. That the range, or difference between the greatest and least average derived from successive groups of small numbers of facts is very considerable, but that it diminishes rapidly as the number of facts increases.

2. That the rate of approximation of the extreme values varies with each different order of facts, and that it does not appear to be amenable to any numerical law.

3. That the greater the number of the elements which determine the occurrence of the events that, when thrown into groups, constitute the materials of our averages, the greater should be the number of our facts; *e. g.* the average duration of life of the entire middle class ought to be deduced from a larger number of facts than the average duration of life of a single profession.

4. That though the *possible* error to which a given small number of facts is liable is very large, there is always a fair probability in favour of any particular average coinciding with, or approaching very closely to, the true average.

5. That the formulæ of the mathematician have a very limited application to the results of observation; and that if incautiously applied, they may lead to very grave errors.

6. That though averages derived from large numbers of facts are worthy of much greater confidence than those founded upon small numbers of facts, the latter class of averages are by no means to be rejected as useless, but should be employed as probabilities of greater or less value, as the number of facts is larger or smaller.

These propositions lay no claim to originality. They are merely intended to express the conclusions legitimately to be drawn from the facts contained in this essay.

Names of the Places from which the Returns in Table X. are taken.

(The numbers are those in column I. of the Table.)

1. England. 2. North-Western Counties. 3. Western Counties. 4. Metropolis. 5. York. 6. South-Western Counties. 7. South-Eastern Counties. 8. South Midland Counties. 9. North Midland Counties. 10. Eastern Counties. 11. Northern Counties. 12. South Wales. 13. Staffordshire. 14. Kent. 15. Gloucestershire. 16. Suffolk. 17. Derbyshire. 18. Cambridgeshire. 19. Middlesex (part of). 20. Huntingdonshire. 21. Alresford and Petersfield. 22. St. George's in the East, London. 23. Northampton. 24. Staines and Uxbridge. 25. Rutlandshire. 26. Winchester. 27. Penkridge. 28. Canterbury. 29. Salisbury.

TABLE X.

Total Births in Seven Years in England, 3,636,383. *Male Births in* 1,000,000, 512,504.

No.	Number of Recorded Births. (Average of Six Years.)	Males Born in 1,000,000 Births. (By Observation.) Average of Six Years, 1839–44.	Males Born in 1,000,000 Births. (By Observation.) Maximum, Minimum, and Range in Six Years, 1839–44.	The same by Calculation. (Formula I.)*	Limit of Error, in excess or defect.	The same by Calculation. (Formula II.)†	Limit of Error, in excess or defect.
			Max. 514,809	500,624	....	513,528	....
1	515,478	512,904	Min. 511,781	499,376	000,624	512,280	000,624
			Range 3,028	1,248	....	1,248	....
			Max. 515,985	505,099	....	518,242	....
2	76,721	513,242	Min. 510,557	494,901	005,100	508,142	005,100
			Range 5,428	10,198	....	10,200	....
			Max. 515,808	505,657	....	518,300	....
3	61,903	512,643	Min. 507,642	494,343	005,657	506,986	005,657
			Range 8,166	11,314	....	11,314	....
			Max. 513,812	505,831	....	517,320	....
4	59,422	511,489	Min. 509,491	494,169	005,831	505,658	005,831
			Range 4,321	11,662	....	11,662	....
			Max. 515,244	506,083	....	518,458	....
5	53,825	512,375	Min. 510,453	493,917	006,083	506,292	006,083
			Range 4,791	12,166	....	12,166	....
			Max. 514,210	506,164	....	517,989	....
6	52,656	511,825	Min. 509,368	493,836	006,164	505,661	006,164
			Range 4,842	12,328	....	12,328	....
			Max. 519,268	506,708	....	518,793	....
7	44,290	512,085	Min. 508,978	493,292	006,708	505,377	006,708
			Range 10,290	13,416	....	13,416	....
			Max. 515,455	507,280	....	517,752	....
8	37,764	510,472	Min. 507,205	492,720	007,280	503,192	007,280
			Range 8,250	14,560	....	14,560	....
			Max. 519,801	507,416	....	521,326	....
9	36,451	513,910	Min. 510,778	492,584	007,416	506,494	007,416
			Range 9,023	14,832	....	14,832	....
			Max. 514,626	507,874	....	520,060	....
10	32,212	512,186	Min. 509,765	492,126	007,874	504,312	007,874
			Range 4,855	15,748	....	15,748	....
			Max. 517,735	508,485	....	523,381	....
11	27,916	514,896	Min. 511,330	491,515	008,485	506,411	008,485
			Range 6,405	16,970	....	16,970	....
			Max. 516,753	510,954	....	525,562	....
12	16,727	514,608	Min. 512,006	489,046	010,954	503,654	010,954
			Range 4,747	21,908	....	21,908	....
			Max. 516,439	511,269	....	523,450	....
13	15,771	512,181	Min. 507,767	488,731	011,269	500,912	011,269
			Range 8,672	22,538	....	22,538	....
			Max. 516,997	511,832	....	523,310	....
14	14,256	511,478	Min. 506,651	488,168	011,832	499,646	011,832
			Range 10,346	23,664	....	23,664	....
			Max. 522,582	512,961	....	525,776	....
15	11,894	512,815	Min. 497,945	487,039	012,961	499,854	012,961
			Range 24,637	25,922	....	25,922	....
			Max. 519,205	514,107	....	526,437	....
16	10,011	512,330	Min. 505,996	485,893	014,107	498,223	014,107
			Range 13,209	28,214	....	28,214	....

TABLE X.—*Continued.*

No.	Number of Recorded Births. (Average of Six Years.)	Males Born in 1,000,000 Births. (By Observation.) Average of Six Years, 1839-44.	Maximum, Minimum, and Range in Six Years, 1839-44.	The same by Calculation. (Formula I.)*	Limit of Error, in excess or defect.	The same by Calculation. (Formula II.)†	Limit of Error, in excess or defect.
			Max. 517,746	516,062	....	528,396	....
17	7,760	512,334	Min. 508,166	483,938	016,062	496,272	016,062
			Range 9,582	32,124	....	32,124	....
			Max. 522,690	518,330	....	530,664	....
18	5,940	514,275	Min. 507,188	481,670	018,330	494,004	018,330
			Range 15,502	36,660	....	36,660	....
			Max. 517,260	522,781	....	531,569	....
19	3,854	508,788	Min. 496,915	477,219	022,781	486,007	022,781
			Range 20,345	45,562	....	45,562	....
			Max. 519,824	531,369	....	534,641	....
20	2,032	503,272	Min. 485,030	468,631	031,369	471,903	031,369
			Range 34,794	62,738	....	62,738	....
			Max. 541,256	535,228	....	550,786	....
21	1,612	515,587	Min. 500,000	464,772	035,228	480,388	035,199
			Range 41,256	70,456	....	70,398	....
			Max. 518,619	536,332	....	542,497	....
22	1,515	506,165	Min. 480,771	463,668	036,332	469,833	036,332
			Range 37,848	72,664	....	72,664	....
			Max. 524,281	543,715	....	549,897	....
23	1,047	506,193	Min. 495,274	456,285	043,715	462,489	043,704
			Range 29,007	87,430	....	87,408	....
			Max. 532,184	548,280	....	565,827	....
24	858	517,578	Min. 498,195	451,720	048,280	469,329	048,249
			Range 33,998	96,560	....	96,498	....
			Max. 527,246	552,631	....	557,122	....
25	722	504,491	Min. 466,942	447,369	052,631	451,860	052,631
			Range 60,304	105,262	....	105,262	....
			Max. 545,031	557,026	....	573,440	....
26	615	516,449	Min. 492,228	442,974	057,026	459,458	056,991
			Range 52,803	114,052	....	113,982	....
			Max. 554,113	567,727	....	583,211	....
27	436	515,557	Min. 493,333	432,273	067,727	447,903	067,654
			Range 60,780	135,454	....	135,308	....
			Max. 556,180	573,722	....	593,267	....
28	368	519,599	Min. 482,500	426,278	073,722	445,931	073,688
			Range 73,680	147,444	....	147,336	....
			Max. 543,798	585,907	....	595,711	....
29	271	509,833	Min. 436,782	414,093	085,907	423,955	085,878
			Range 107,016	171,814	....	171,756	....

* Formula I. is based on the assumption that the chances of two events (in this case, a male or female birth) are equal. The formula is $\sqrt{\frac{2}{\mu}}$ where $\mu =$ the total average births for one year, as given in the second column of the Table.

† Formula II. is the well-known formula for calculating the limits of error to which any number of observations on two alternative events are liable. This formula is $\frac{m}{\mu} \pm 2\sqrt{\frac{2 \cdot m \cdot n}{\mu^3}}$ where $m =$ the total of one of the two events m and n, and μ their sum, or $m + n$.